AI-ДАЙДЖЕСТ

Anthropic зсередини: 15% часу на код

Bun-рерайт зашипів і крутить Claude Code просто зараз, показує Ґерґелі Орош.

середа, 29 липня 2026

У ЦЬОМУ ВИПУСКУ

1. Всередині Anthropic: 15% часу на код, 85% на те, щоб він працював The Pragmatic Engineer, 28.07. Орош фізично сходив у їхній офіс і поговорив з чотирма інженерами, серед них Джарред Самнер, автор Bun, і Кейтлін Лессе, голова інженерії Claude Platform.

Цифри, які варто повісити на стіну:

• **Реалізація рерайту Bun зайняла ~15% часу, решта 85% - компіляція, тести, перевірка.** Дослівно Таріка Шіхіпара: «**дуже мало токенів іде на саму реалізацію; більшість - на дослідження невідомого, прототипування, макетування, а потім на верифікацію і тестування**»

• **64 паралельних агенти**, $165к токенів по API-прайсу, **535 496 рядків Zig**. Самнер прямо каже, що вручну це були б «три інженери з повним контекстом на рік»

• Кожен інженер лабораторії ганяє **3-10 агентів одночасно**, ліміту токенів нема взагалі • Максимум **два інженери на проєкт**

Чому це важливо, і це поправка до вчорашнього. Вчора в пункті 3 був розбір Локвуда: релізу нема 11 тижнів, 2 475 відкритих PR, реальна ціна ближча до $800к. Це лишається правдою про **публічний репозиторій Bun**. Але сьогодні з першоджерела видно те, чого в тому розборі не було: **рерайт змерджено і він у продакшені, на ньому працює Claude Code**. «Зроблено» і «черга PR розсмокталась» - дві різні речі, і Локвуд міряв другу.

Практичний висновок той самий, тільки тепер з цифрою: якщо 85% роботи - це верифікація, то цінність проєкту з агентами впирається в те, чи є чим перевірити результат.

Окремо: **вони видалили 80% системного промпту Claude Code**, бо «модель порозумнішала». Та сама цифра з гайду 25.07, але тепер сказана інженером зсередини як рутина.

2. Claude знайшов математичні дірки в криптографії, і це вже не пошук багів у коді 192 бали HN. Anthropic Frontier Red Team, 28.07. Найсерйозніший технічний результат доби.

Два результати з першоджерела:

• **HAWK** - кандидат NIST на постквантовий підпис, який пройшов **два роки експертного розгляду**. Mythos Preview знайшов невикористану симетрію (нетривіальний автоморфізм) у ґратці за **60 годин** і фактично **вдвічі порізав ефективну довжину ключа**. Щоб компенсувати, ключі треба подвоїти, а це вбиває сенс схеми як кандидата • **AES з урізаною кількістю раундів** - прибрано одну з здогадок атакуючого, швидкість найкращої відомої атаки зросла **в 200-800 разів**

Чесні мітки, які вони ставлять самі: атака на AES не переноситься на продакшн, це ослаблена версія, повний шифр не зламаний. Вартість: **~$100 000 API-косту на кожен результат**. AES-атаку Claude знайшов **повністю автономно** у скафолді, HAWK - у парі з дослідником.

Чому це важливо. Ключова деталь в тому, ЯК це зроблено: «Claude Code-подібний харнес з кількома робочими агентами в пісочниці». Знову те саме слово, що йде через увесь тиждень, **харнес**. І ще: модель тут знайшла ваду в **самому алгоритмі**, який два роки дивились люди. Це інший клас задачі, ніж перевірка коду.

3. MCP переїхав на stateless, і це ламає сумісність 110 балів. Специфікація **2026-07-28**, вийшла вчора.

Що змінилось, з блогу мейнтейнерів:

• **Хендшейк і сесії скасовано** - `initialize` / `initialized` і хедер `Mcp-Session-Id` **офіційно на пенсії**. Кожен запит самоописовий, тож будь-який може лягти на будь-який інстанс за звичайним round-robin балансувальником • Імена методів і тулів їдуть у хедерах `Mcp-Method` і `Mcp-Name`, шлюзи можуть роутити й авторизувати, не заглядаючи в тіло • Запити «сервер → клієнт» (sampling, elicitation) переїхали на **MRTR**, постійно відкриті двонаправлені стріми більше не потрібні • **Списки тулів кешовані** з детермінованим порядком, щоб апстрімний prompt cache не інвалідувався при реконектах • Відхід від Dynamic Client Registration до **CIMD**, і формальна політика депрекації з вікном **мінімум 12 місяців**

Масштаб для контексту: **близько пів мільярда завантажень на місяць** по Tier-1 SDK, TypeScript і Python перетнули мільярд сумарно.

Чому це важливо. Кешовані списки тулів з детермінованим порядком означають менше промахів по кешу там, де тули підвантажуються на вимогу. Оновлювати нічого не треба, сервери підтягнуться, коли їх оновлять автори. А згадка `Mcp-Session-Id` у лозі тепер означає депрекацію.

4. OpenAI викотила Codex Security - сканер вразливостей як CLI 404 бали, 126 коментарів. `@openai/codex-security` - CLI і TypeScript SDK: сканує репозиторії, рев'ює зміни, тримає історію знахідок і ходить у CI.

З README: `npm install @openai/codex-security`, потім `login` і `scan.`. Для CI - `OPENAI_API_KEY`. Потребує Node 22+ і Python 3.10+.

Чому це важливо. Пряма паралель з пунктом 1: Самнер розповідає, що для Rust-рерайту вони прогнали **11 запусків Claude Security Scanner**, плюс фаз-тестування. Обидві лабораторії зійшлись на тому, що безпекове сканування має бути окремим автоматом у конвеєрі. Разового рев'ю мало. Для будь-якого репозиторію, де лежать токени і службові ключі, це найдешевша страховка.

5. Файнтюн 9B-моделі за $500 обійшов фронтір, у 40 разів дешевше 311 балів. Fermisense, задача - перевірка товарного каталогу.

Цифри з першоджерела: GRPO-файнтюн **9B опенсорсної моделі** б'є **кожну** протестовану фронтір-конфігурацію на тій самій задачі, з тими самими тулами, картинками і скорером. Ціна: **$0.50 за 1000 лістингів**, це **у 40 разів дешевше за найдешевшу фронтір-збірку і ~340× за найдорожчу**.

Чому це важливо. Третій незалежний доказ патерну тижня (після MAI-Cyber-1-Flash учора і харнесу Meta в понеділок): **на вузькій, добре визначеній задачі дешева спеціалізована модель б'є універсальну дорогу**. Рамка для продукту проста: якщо всередині є задача, яку роблять тисячами однакових ітерацій, її не треба віддавати фронтіру.

6. Kimi K3 розібрали по архітектурі, і виявилось, що це стара лінія 346 балів. Себастьян Рашка, 28.07, продовження вчорашнього пункту 2.

Головне: K3 - це **масштабована продакшн-версія Kimi Linear**, яку виклали ще торік. Ріст **48B → 2.8T**, і це зараз **найбільша відкрита модель у світі**. Єдиний справді новий компонент - **LatentMoE** (той самий, що в Nemotron 3 Ultra): стискає великі лінійні шари так само, як multi-head latent attention стискає увагу.

Загальний вектор, спільний з Nemotron 3 і DeepSeek V4 - **«дешевше в інференсі»**: MoE → LatentMoE, звичайна увага → MLA і Kimi Delta Attention. Єдина зміна поза ефективністю - **attention residuals**: зв'язують залишкові шляхи між шарами через attention-скор.

Чому це важливо. Вчора K3 подали як «китайці догнали позаминулий фронтір». Сьогоднішній розбір уточнює: ривка навздогін не було, **два роки шліфувалась одна лінія**, масштабована в 58 разів. Це важливіше за бенчмарки: наступна ітерація

прилетить не з нуля. І вже сьогодні на HN є **запуск K3 на M1 Max** (111 балів), тобто відстань від релізу ваг до «крутиться на ноуті» тепер вимірюється днями.

7. levelsio ловить падіння виторгу і сам себе діагностує Найгучніша дискусія за добу: 882K переглядів на головному пості, 225K на продовженні.

Дослівно: «**Бачу тренд падіння виторгу і трафіку в інді-хакерів. На своїх проєктах теж… Здається очевидним, що BigAI канібалізує все, що раніше було застосунками**». І окремо, з іронією до себе: «**Скасовані і вайбкодені 100% SaaS-підписок. Оплата лишилась тільки за домени, хостинг, сховище і AI API. Іронія, що самого замінили після того, як замінено все, за що платилось, не втрачена**».

Найсильніший контраргумент у треді, від Віка: «**продаж сервісним бізнесам - сантехнікам, чистильникам басейнів. Припущення, що в них з'явиться бажання і час вайбкодити заміну - це ілюзія**». levelsio погодився: складні бізнеси, продані нетехнічним IRL-компаніям, тримаються.

Чому це важливо. Вчора Леві казав «корпорати наймають, змінився профіль», позавчора Стенфорд - «в агрегаті ефект малий». Сьогодні levelsio додає четвертий замір, і картина складається. **Б'є по тому, що легко відтворити за $9/міс. Не дістає туди, де потрібна доменна робота з живими людьми.**

8. «Не варто просити у LLM оцінку впевненості» - стаття, варта прочитання 87 балів. Джастін Флік, 27.07.

Теза жорстка і дослівна: просити модель видати число, наскільки вона впевнена у власній відповіді, «**за всім видимим, абсолютно марно**». Найбільше дратує саме **безперервна шкала 0-100**: це «психологічний трюк безпеки - робить вивід відчутно надійнішим, не роблячи його надійнішим», і ті, хто це ставить, «здебільшого брешуть самі собі».

Чому це важливо. Критика б'є по шкалі, яку модель виставляє сама собі. Мітки на кшталт [proven] / [promising] / [fuzzy] влаштовані інакше: кілька дискретних категорій, прив'язаних до **зовнішньої якості доказів** (є дослідження / є сигнал / це здогадка). Така мітка описує стан літератури, а самовідчуття моделі до неї не входить. Критику вона витримує, але тільки поки її ставлять чесно. [proven] без назви джерела - це вже той самий трюк.

9. Наваль ставить незручне питання про відкриті ваги 282K переглядів, 262 відповіді, найгучніший одиночний пост доби. Закриває сюжет, що ведеться з 25.07.

Дослівно: «**Якби у відкритих вагах були заховані бекдори і упередження, можна не сумніватись, що лабораторії із закритими вагами їх знайдуть і оприлюднять**».

Чому це важливо. Одне речення перевертає аргумент про небезпеку відкритих ваг у аргумент **на їхню користь**. Відкриті ваги перевіряються конкурентами, у яких є і мотив, і ресурс копати. Закриті - ніким. Разом з цитатою Дженсена з понеділка це закриває суперечку з двох боків.

10. uv 0.12.0 - перший мажор з березня, і він ламає `uv init` 120 балів. Astral, 28.07.

Головна поламка: **`uv init` тепер за замовчуванням оголошує build-систему** (`uv_build`), кладе код у `src/<name>` і додає `[project.scripts]`. Раніше створювався непакетований проєкт з `main.py` без build-системи. **Наявні проєкти не зачеплені**.

Одна дія: якщо в `[build-system]` стоїть верхня межа на `uv_build`, її треба підняти до `uv_build>=0.11.32,<0.13`.

Чому це важливо: нічого термінового. Але при наступному створенні проєкту `uv init` поводитиметься інакше, ніж очікується.

Misc • **Позиція Anthropic по відкритих вагах добрала 1153 бали і 1691 коментар** - учорашній пункт 1 став топ-темою доби на HN, але сам текст не змінився • **7.1 землетрус у Японії** (780) - найвищий небізнесовий топ доби • **Нова вакцина від ВІЛ показала безпрецедентний успіх** у доклінічному дослідженні (597, La Jolla Institute)

• **Європейська громадянська ініціатива проти цифрового ID і вікової верифікації** (544) - «Stop Killing the Internet»

• **«Substack-письменникам потрібен власний сайт»** (459)

• **Netflix звільнив працівника** за те, що той поділився особистим на тренінгу довіри (424, 479 коментарів)

• **SlopCodeBench добрав до 390 балів** - учорашній пункт 4, без нових даних • **Пропущений підкреслювач і 18 місяців в'язниці** (385) - вчора був у misc, сьогодні добрав коментарів • **Kimi Linear** - та сама стаття 2025 року, з якої виріс K3 (295), і **розбір сімейства DeltaNet** (286)

• **GrapheneOS** знову в топі двічі (238 + 154) - тепер про блокування обшуку на кордоні • **Zig: внутрішня кухня інкрементальної компіляції** (209)

• **macOS Tahoe 26.6** - безпекові оновлення (200)

• **DMARC публічний з 2012, але більшість доменів компаній його не вмикають** (178)

• **Half-Life портували на Mac OS 9** (186) - просто тому що змогли • **Kimi K3 крутиться на M1 Max** (111) і вже доступний через Telnyx API (129)

• **Astronauts описують стійке відчуття «спостерігача»** після шестимісячних місій (182)

• **Втрата кисню в океані загрожує стабільності планети** (141, Scripps)
